



Comparative Lesion Sequencing Provides Insights into Tumor Evolution

Citation

Jones, Siân, Wei-dong Chen, Giovanni Parmigiani, Frank Diehl, Niko Beerenwinkel, Tibor Antal, Arne Traulsen, et al. 2008. Comparative lesion sequencing provides insights into tumor evolution. *Proceedings of the National Academy of Sciences* 105(11): 4283-4288.

Published Version

doi:10.1073/pnas.0712345105

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:11644153>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

Comparative lesion sequencing provides insights into tumor evolution

Siân Jones^{*†}, Wei-dong Chen^{†§¶†}, Giovanni Parmigiani^{*}, Frank Diehl^{*}, Niko Beerenwinkel^{||**}, Tibor Antal^{**}, Arne Traulsen^{**&}, Martin A. Nowak^{**}, Christopher Siegel^{††}, Victor E. Velculescu^{*}, Kenneth W. Kinzler^{*}, Bert Vogelstein^{*§§}, Joseph Willis^{††§§§}, and Sanford D. Markowitz^{†§¶ §§}

^{*}The Ludwig Center for Cancer Genetics and Therapeutics, Department of Biostatistics, Howard Hughes Medical Institute, and Sidney Kimmel Cancer Center at the Johns Hopkins Medical Institutions, Baltimore, MD 21231; Departments of [†]Medicine, ^{††}Pathology, ^{††}Surgery, and [§]Ireland Cancer Center, at Case Western Reserve University and Case Medical Center of University Hospitals of Cleveland, and [¶]Howard Hughes Medical Institute, Cleveland, OH 44106; [&]Department of Biosystems Science and Engineering, ETH Zurich, Basel Switzerland; ^{||}Max-Planck Institute for Evolutionary Biology, Ploen, Germany; ^{**}Program for Evolutionary Dynamics, Harvard University, Cambridge, MA 02138

[†]S. J. And W. C. contributed equally to this work.

^{§§}To whom correspondence should be addressed. E-mail: vogelbe@jhmi.edu (B.V.); JosephE.Willis@UHhospitals.org (J. W.); sxm10@cwru.edu (S.D.M.)

Classification: Biological Sciences-Medical Sciences

Key Words: Colorectal Cancer, Metastasis, Tumor Evolution; Cancer Genetics; Stem Cells

Manuscript Includes: 12 pages of text, 5 figures, 1 table and supplementary information with 1 table

Abbreviations: CI, 90% confidence interval.

We show that the times separating the birth of benign, invasive, and metastatic tumor cells can be determined by analysis of the mutations they have in common. When combined with prior clinical observations, these analyses suggest the following general conclusions about colorectal tumorigenesis: (i) It takes ~17 years for a large benign tumor to evolve into an advanced cancer but less than two years for cells within that cancer to acquire the ability to metastasize; (ii) it requires few, if any, selective events to transform a highly invasive cancer cell into one with the capacity to metastasize; (iii) the process of cell culture *ex vivo* does not introduce new clonal mutations into colorectal tumor cell populations; and (iv) the rates at which point mutations develop in advanced cancers are similar to that of normal cells. These results have important implications for understanding human tumor pathogenesis, particularly those associated with metastasis.

Colorectal tumorigenesis proceeds through well-defined clinical stages associated with characteristic mutations (1, 2) (Fig. 1). The process is initiated when a single colorectal epithelial cell acquires a mutation in a gene inactivating the *APC/β-catenin* pathway (1). Mutations that constitutively activate the *KRAS/BRAF* pathway are associated with the growth of a small adenoma to a clinically significant size (>1 cm in diameter) (3). Subsequent waves of clonal expansion driven by mutations in genes controlling the TGF-β (4, 5), *PIK3CA* (6), *p53* (7) and other pathways are responsible for the transition from a benign tumor (adenoma) to a malignant tumor (carcinoma). The only difference between a carcinoma and an adenoma is the ability of the former to invade the tissues underlying the colorectal epithelium. Some tumors eventually acquire the ability to migrate and seed other organs (metastasis) (8). Colorectal tumors can usually be cured by surgical excision at any stage prior to this last one, i.e., prior to metastasis to distant sites such as the liver (9).

Understanding the basic features of this evolutionary process has obvious and important implications for both scientific and medical research. But many questions remain. For example, how long does it take for a particular neoplastic cell to acquire the genetic events required for each sequential step in this progression? This question has heretofore been impossible to address in individual patients, though relevant information about bulk tumors, rather than cells, has been obtained through clinical and radiographic studies (10-12). We here describe an approach that can answer this and related questions.

Large scale sequencing of the vast majority of protein coding genes in human tumors has recently become possible and was applied to study the genomes of breast and colorectal cancers (13, 14). In the current study, we investigated whether the mutations discovered in the colorectal cancers evaluated in Wood *et al.* (14) were found in other neoplastic lesions from the same patients, an approach we call "comparative lesion sequencing". We show that the sequencing data, when analyzed quantitatively, can be used to determine the time intervals required for development of the cells responsible for any two sequential clonal expansions. We were particularly interested in the expansions associated with metastasis. This final expansion is the least well understood at the biochemical and physiologic levels, even though it is responsible for virtually all deaths from the disease.

Results

Point Mutation Rates and Growth Kinetics of Colorectal Cancers. Though knowledge of the precise mutation rate and tumor growth rates of these lesions are not required to make conclusions from comparative lesion sequencing, estimates of these parameters can inform their interpretation. An estimate of the point mutation rate in these tumors can be made on the basis of the results reported in (14) wherein 847 non-synonymous mutations were detected among 304 million bp sequenced at high quality. All of these mutations were somatic, i.e., not present in the germ-line. Most of the lesions evaluated in (14) were liver metastases and all were mismatch repair proficient. In order to convert the mutation *prevalence* found in Wood *et al.* to a mutation *rate*, it is necessary to know the number of divisions that the cancer cell had undergone. The most reliable way to measure cell division time in human tumors is through the administration of DNA precursors such as BrdU to patients, followed by evaluation of BrdU incorporation plus DNA content via flow cytometry (15). This approach yields T_{pot} , defined as the time between cell divisions in the absence of cell death. Several hundred colorectal cancers have been evaluated by this method, with T_{pot} measured as ~ 4 days (16-21). Using this figure for the cell division rate and the mutational data reported in (14), the point mutation rate in colorectal cancers is estimated to be 4.6×10^{-10} mutations/bp/generation. This rate is slightly less than that measured in various normal cell types ($\sim 10 \times 10^{-10}$ mutations/bp/generation; (22-25)). Additional details about this estimate are provided in the Supplementary Information.

Comparison of mutations before and after cell culture or xenografting. The samples used in (14) were all derived from colorectal cancer cells that had been passaged for six to twelve months *in vitro* as cell lines or in nude mice as xenografts. Before initiating the current study of other lesions from the same patients, it was important to determine whether the mutations identified in the cultured or xenografted cells were actually present in the naturally occurring lesions prior to cellular expansion *ex vivo*. For this purpose, we analyzed 289 different mutations found in 18 cell lines or xenografts, each derived from a different patient. Five of these had been initiated and passaged *in vitro*, while the remaining 13 had been passaged as xenografts in nude mice. Two hundred eighty-seven of the 289 mutations (99.3%) found in the cell lines or xenografts were also found in the original tumors. The direct Sanger sequencing method we used in the experiments reported herein had a sensitivity of $\sim 25\%$, so that a heterozygous mutation present in less than 50% of the cells would not be observed (Fig. 2). These data therefore indicate that the point mutations found in colorectal cancer cell lines or xenografts only rarely arise during *in vitro* or *in vivo* experimental growth of cells after the tumors are excised from patients.

Comparison of Metastases with Primary Colorectal Cancers. Paired samples of primary colorectal cancers and metastatic lesions from ten patients were available for this study. Of the index lesions (Table 1), seven had been excised from the liver and three from mesenteric lymph nodes. We were able to evaluate an average of 28 mutations per lesion in the patients evaluated in the Discovery Screen of (14). The remaining index lesions had been studied only in the Validation Screen, so only ~ 5 mutations could be studied in these cases. In all, 233 somatic mutations identified in the index metastases were evaluated in the ten cases. Of these, only 7 (3.0%, CI 1.5 to 5.7%) were *not* found in the colorectal cancers from which the metastases arose (SI Table 2).

Comparisons of Different Metastatic Lesions. In 11 of the 13 patients, we were able to examine at least one metastatic lesion different from the index lesion of the same patient. Of a total of 261 mutations evaluated, 255 (97.7%) were found in the 29 additional metastases

studied (Table 1). This was the expected result, as the great majority of the mutations present in metastases were also present in the precursor advanced colorectal carcinoma, as noted above. What was more informative was the study of patients in whom mutations were identified in the metastasis but not in the precursor advanced carcinoma (henceforth denoted "metastasis-specific mutations"). Though there were only seven of these mutations identified, five of them were particularly informative as they could be assessed in other metastatic lesions from the same patients. In patient 5, both of the metastasis-specific mutations originally identified in the index liver metastasis were also identified in a mesenteric lymph node metastasis. In patient 7, three liver metastasis-specific mutations identified in the index metastasis were identified in a second, independent liver metastasis concurrently excised from the patient.

Comparison of Advanced Colorectal Carcinomas with Large Adenomas. It is believed that colorectal carcinomas arise in pre-existing, benign adenomas through the acquisition of additional genetic alterations (Fig. 1). In most cases, the adenomatous tissue is destroyed during carcinoma growth. However, in two of the cases studied here, large residual adenomas at the edge of the carcinomas were still present (Fig. 3). Paraffin-embedded sections of these lesions were carefully microdissected to separate adenomatous from carcinomatous elements and evaluated for 33 mutations known to be present in the carcinoma. Ten of the 33 mutations (30%) were not found in the adenomatous components. The differences between the fraction of mutations found in metastases but not their precursor advanced carcinomas and the fraction of mutations found in the advanced carcinomas but not in their large precursor adenomas were statistically significant ($P < .001$, two-group binomial test for equality of proportion).

Of the ten mutations identified in advanced carcinomas but not large adenomas, seven were in candidate cancer genes (CAN-genes) as defined in Wood *et al.* (14). This proportion is significantly different than the proportion of metastasis-specific mutations (12.5%) that were in CAN-genes ($P < .001$, two-group binomial test) or the proportion of mutations that were in CAN-genes among all genes with mutations (16%, $p < 0.01$, two-group binomial test). Two of the ten mutations identified in carcinomas but not in their precursor adenomas were in *TP53*, consistent with prior data on the timing of *TP53* mutations (26).

Quantification of the Level of Mutations in DNA. The absence of a somatic mutation in a given DNA sample, as assessed by Sanger sequencing, simply means that the mutation was not present in $>25\%$ of the analyzed DNA template molecules (i.e., $>50\%$ of the cells in the case of heterozygous mutations). Mutations present in a smaller fraction can generally not be distinguished from the background in sequencing chromatograms. To determine whether the mutations were present in a smaller but still sizable fraction of the tumor cell population, we evaluated a subset of the DNA samples via BEAMing (beads, emulsions, amplifications, magnetics) assay (see SI Methods) (27, 28). We performed 20 BEAMing assays in 7 patients, focusing on those mutations that appeared to be present in late stage lesions but not in an earlier stage lesion of the same patient (e.g., present in metastasis but not in the advanced colorectal carcinoma). In 19 of these assays, no mutations were observed (examples in Fig. 4). As the sensitivity of the BEAMing assays was $\sim 0.01\%$, we conclude that less than 1 in 2500 cells in the precursor lesion contained any of these 19 mutations, thus suggesting that at least one major clonal expansion occurred between the two stages analyzed in each case.

Colorectal Cancer Evolution: Mathematical Assessment. The data in Table 1 can be used to determine the relative timing of the birth of the founder cells (Fcells) that gave rise to the various tumor cell populations described above (Fig. 5). The basis for this analysis is that all somatic mutations present in clonal fashion in an adenoma (i.e., present in all cells of the tumor) must have been present in its cell of origin (its founder cell). These mutations accumulated during the life span of this founder cell and include those that occurred during the turnover of normal stem cells prior to the onset of tumorigenesis. As tumors progress, they accumulate additional mutations which become fixed in the founder cells of subsequent neoplastic states. The founder cell of the advanced carcinoma, for example, will harbor all the mutations present in the precursor adenoma plus additional mutations that occurred in the interim. The length of this interim period can be estimated by measuring the number of additional mutations in the progressed lesion.

The founder cells of interest are (i) the one ($F_{cell_{Met}}$) that gave rise to the final clonal expansion resulting in the index metastasis; (ii) the last common ancestor ($F_{cell_{ACa}}$) of the advanced carcinoma and $F_{cell_{Met}}$; and (iii) the last common ancestor ($F_{cell_{LAd}}$) of the large adenoma and $F_{cell_{ACa}}$. The birth date of a founder cell is defined as the age of the patient when the founder cell underwent its first division. As shown in the SI *Text*, the interval ($\Delta T_{ACa, Met}$) between the birth date of founder cells $F_{cell_{Met}}$ and $F_{cell_{ACa}}$ is

$$\Delta T_{ACa, Met} = F_{ACa, Met} \cdot T_{Met} \quad [1]$$

where $F_{ACa, Met}$ is the fraction of the mutations in the metastasis that were *not* found in the advanced carcinoma (i.e., $1 - [\text{number of mutations in advanced carcinoma} / \text{number of mutations in metastasis}]$). Similarly, the interval ($\Delta T_{LAd, ACa}$) between the birth dates T_{LAd} and T_{ACa} of founder cells $F_{cell_{ACa}}$ and $F_{cell_{LAd}}$, respectively, is

$$(\Delta T_{LAd, ACa}) = F_{LAd, ACa} \cdot T_{ACa} \quad [2]$$

where $F_{LAd, ACa}$ is the fraction of mutations in the advanced carcinoma that were *not* found in the large adenoma. Similar equations can be applied to any two lesions that represent the clonal expansions of two founder cells as long as one of the two founder cells is a direct descendent of the other.

Note that these equations are entirely independent of the actual mutation rates and cell division times (T_{pot}), which likely vary among different patients and cancers. Additionally, they are independent of whether the observed mutations are "drivers" or "passengers", i.e., whether the mutations causally contribute to tumor progression. They only require that the mutation rate and cell division times, whatever they are, are constant throughout each patient's life. As described in the SI, this requirement is, in general, expected to be met. Mutations thereby act as a clock, providing information similar to that obtained through the use of sequence divergence to assess the relatedness of organisms or cells during evolution or development (29, 30).

Application to individual patients. One of the major results of the current study is that $F_{LAd, ACa}$ is much greater than $F_{ACa, Met}$, meaning that it takes much longer for a large adenoma to evolve into an advanced carcinoma than for such a carcinoma to metastasize. Assumptions that limit the accuracy of the times determined through these equations are given in the SI *Text*. Their implementation can best be illustrated through their application to five patients in the current study in whom a minimum of 25 mutations could be evaluated (Table 1).

Patient 1 was 73 years old when she developed an advanced carcinoma of the ascending colon that was 4 cm in diameter and of stage T3N1M0 (T3 refers to the stage of the carcinoma, which in this case had grown completely through the muscularis propria; N1

indicates that cancer was found in at least one but less than four lymph nodes; M0 indicates that no distant metastases were evident at the time of surgery). Fifteen months later, a liver metastasis of 5 cm in diameter was found to have developed. All 47 mutations found in the metastasis were also found in the advanced carcinoma in the colon ($F_{ACa, Met} = 0.0$). Application of [1] indicated that the metastasis originated from a cell ($F_{cell_{Met}}$) whose birth occurred very soon after the birth of the cell ($F_{cell_{ACa}}$) that founded the advanced carcinoma (CI 0 to 3.3 years).

Patient 3 was 83 years old when she developed an advanced carcinoma of the ascending colon that was 9 cm in diameter and of stage T4N2M1 (N2 indicates that cancer cells were found in more than 3 mesenteric lymph nodes). A residual adenoma that surrounded the carcinoma was identified at the time of surgery. A small (< 1 cm diameter) mesenteric lymph node metastasis was found to contain 25 mutations that were subsequently evaluated in other lesions of this patient. Of these, 24 were found in the colorectal carcinoma ($F_{ACa, Met} = 0.04$). Application of [1] indicated that the advanced carcinoma founder cell was born 3.2 years (CI 0.4 to 7.1 years) before the lymph node metastasis founder cell was born. In contrast, evaluation of the same mutations in the large adenoma from which the carcinoma developed revealed an $F_{LAd, ACa}$ of 0.23. Application of [2] indicated that the large adenoma founder cell was born 17 years (CI 7.7 to 30.6 years) before the advanced carcinoma founder cell. In the ~17 years between the birth of $F_{cell_{LAd}}$ and $F_{cell_{ACa}}$, the tumor underwent waves of clonal expansion driven by mutations in *TP53* and the other genes (SI Table 2) presumably required for invasion and further growth of this tumor. Once it acquired these capabilities, a cell ($F_{cell_{Met}}$) capable of lymph node metastasis appeared within a relatively short period of time.

Patient 5 was 72 years old when he developed an advanced carcinoma of the sigmoid colon that was 1.5 cm in diameter and of stage T3N2M1, accompanied by an 8.9 cm liver metastasis. Comparative lesion sequencing indicated that the metastasis founder cell $F_{cell_{Met}}$ was born 2.8 years (CI 0.6 to 5.0 years) after the birth of the advanced carcinoma founder cell $F_{cell_{ACa}}$. A large (1.3 cm) mesenteric lymph node metastasis and two smaller mesenteric lymph node metastases were also evaluated from this patient. The larger lymph node contained the same 50 mutations identified in the liver metastasis, including the two mutations not found in the primary colorectal carcinoma; the two smaller lymph nodes did not contain these two mutations. Thus the 1.3 mesenteric lymph node metastasis and liver metastasis founder cells may have both been derived from a small population of cells within the carcinoma that had acquired metastatic capability. Alternatively, the liver metastasis could have originated from the large mesenteric lymph node metastasis. In this case, comparative lesion sequencing indicates that the liver metastasis founder cell must have been born soon after (0 years, CI 0.0 to 2.0 years) the birth of the lymph node metastasis founder cell.

Patient 7 was 55 years old when she developed an advanced carcinoma of the ascending colon that was 3.5 cm in diameter and stage T3N1M0. Twenty months later, two metastases of 3.5 and 4 cm diameter were found in the liver. Comparative lesion sequencing of the 4 cm liver metastasis and the colorectal cancer indicated the metastasis founder cell was born 6.6 years after the carcinoma founder cell (CI 1.8 to 8.6 years). Two mesenteric lymph node metastases removed at the time of the initial surgery and the 3.5 cm liver metastasis noted above were also evaluated. Three metastasis-specific mutations were identified in both liver metastases but not in either nodal metastasis.

Patient 13 was 55 years old when he developed an advanced carcinoma of the ascending colon that was 2.5 cm in diameter and stage T3N1M1. A 7 cm metastasis in the right lobe of the liver and a metastasis in a mesenteric lymph node were removed at the time

of surgery. Twenty-nine months following this resection, a new liver metastasis of 3.1 cm diameter was detected in the left lobe and completely excised. One year later, another metastasis in the liver, of diameter 3.5 cm, was identified. The metastases that were identified 29 months and 41 months after the initial diagnosis both had 19 mutations that were not found in the advanced carcinoma or metastatic lesions excised at the initial surgery, with $F_{\text{ACa, Met}} = 0.28$. In contrast, all of the mutations identified in the metastatic lesions removed at the initial surgery were also present in the advanced carcinoma removed concurrently. We interpret this result in the following way. Chemotherapy consisting of irinotecan, leucovorin, and 5-FU administered in the nine months following the initial surgery pruned most of the micrometastatic cells remaining in the liver. One of these cells was resistant to the chemotherapy and became the founder cell of the new metastasis and its later recurrence. The chemotherapy had induced many new mutations in this cell, consistent with the known mutagenicity of irinotecan and perhaps exacerbated by the 5-FU (31). Equation [1] cannot be used to estimate the relative birth date of this cell as comparative lesion sequencing requires that the mutation rate be constant throughout the tumorigenic process (see SI). It is notable that this patient was the only one of the patients analyzed in depth in our study who had been treated with irinotecan prior to the development of a new metastatic lesion.

Discussion

A Temporally-defined Model of Colorectal Cancer. The data and approach used in the current study can be used to temporally model some of the key genetic events in colorectal tumorigenesis. As illustrated in Fig. 5, comparative lesion sequencing suggests that the average time interval between the birth of a large adenoma founder cell and the births of an advanced carcinoma founder cell is 17 years (CI 10.8 to 26.3 years). However, the average interval between the birth of the advanced carcinoma founder cell and the liver metastasis founder cell is only 1.8 years (CI 0.9 to 3.1 years).

Information about the birth times of the founder cells giving rise to various neoplastic stages has not heretofore been available. However, our estimates of these values are consistent with clinical and radiological observations on bulk tumors. For example, the time between the appearance of small adenomas and the diagnosis of a carcinoma has been estimated at 20 to 25 years from studies of patients with familial adenomatous polyposis (11). Similarly, serial studies of sporadic colorectal tumor patients have indicated that the transition from large adenoma to carcinoma takes ~15 years (11). Our estimates are also consistent with the long doubling times of tumors determined by serial radiologic studies or serial measurements of the CEA serum biomarker (10, 12, 32, 33). Such studies have indicated mean doubling times that are generally 2 to 4 months in metastases and much longer in adenomas and carcinomas.

Biological Implications. Our findings suggest that virtually all of the mutations necessary for metastasis are already present in all of the cells of the antecedent carcinoma. These data are compatible with two distinct models. In Model A, none of the carcinoma cells can give rise to a metastasis, but they are close to being able to do so; one or a few more genetic alterations are required. In Model B, all of the carcinoma cells can give rise to metastasis; no more genetic alterations are required. Data derived from the current study, involving comparisons of different metastatic lesions from the same patients, are compatible with either model.

Model A: If every cell in the cancer cell population were capable of giving rise to a metastasis, it is extremely unlikely that any two independent metastases would harbor an identical mutation not found in the carcinoma. However, as described in the Results, we identified five metastasis-specific mutations that were each present in more than one metastasis from the same patient (patients 5 and 7). If the founder cells of one of these two metastases were not a direct descendent of the other, these data would support the idea that a small population of cells within the carcinoma had acquired additional alterations that endowed them with the capacity to metastasize. Such alterations could be the point mutations actually identified as metastasis-specific (SI Table 2) or any other heritable event (whole chromosome gains or losses, chromosome translocations or amplifications, or certain epigenetic changes (34).

Model B: General support for this model comes from the fact that there so few additional alterations identified in the metastases compared to the advanced carcinomas. The finding that mutations not found in the carcinoma were identified in two anatomically distinct metastases could be explained if the founder cells of the two metastases had both come through a bottleneck after they migrated from the primary colorectal carcinoma. In Patient 5, this could have occurred if the liver metastasis had developed from a cell within the mesenteric lymph node metastasis that contained the same mutations. In Patient 7, this could have occurred if both liver metastases' founder cell had developed in a lymph node metastasis that was not detected or excised.

The reason that progression of the large adenoma to advanced carcinoma takes so much longer than the progression of the latter to metastasis is presumably because many more mutations and clonal expansions are required (some of which are indicated in Fig. 5). Moreover, some of the genes responsible for the adenoma to carcinoma progression have been identified (SI Table 2 and Fig. 5). One reasonable interpretation of the data is that the capacity to invade through layers of the bowel wall without dying, thereby becoming an advanced colorectal cancer, is the most challenging step in the process that eventually leads to metastasis. Once that step occurs, few additional steps are required for metastasis to take off. The advent of large-scale cancer genome sequencing provides uniquely valuable biomarkers to study tumor evolution. The study of additional mutations and lesions using the approach described in this work could definitively answer a variety of long-standing questions about the basic nature of the metastatic process in humans (35-39).

Acknowledgements

We thank D. H. Nguyen for preparing Figure 5, J. Lutterbaugh and L. Kasturi for expert technical assistance, Y. He for advice about flow cytometry, and D. Shibata, D. Williams and J.T. Vogelstein for critical comments. This study was supported by Virginia and D. K. Ludwig Fund for Cancer Research, PHS grants CA43460, CA57345, CA62924, CA121113, CA116867, CA120237, CA127306, CA043703, CA62924, CA105090 and GM078986 and by a gift from the National Colon Cancer Research Alliance. N. B. was partially supported by a grant from the Bill & Melinda Gates Foundation through the Grand Challenges in Global Health Initiative.

References

1. Kinzler, K. W. & Vogelstein, B. (1996) Lessons from Hereditary Colon Cancer. *Cell* 87:159-170.
2. Fearon, E. R. & Vogelstein, B. (1990) A genetic model for colorectal tumorigenesis. *Cell* 61:759-67.
3. Vogelstein, B., Fearon, E. R., Hamilton, S. R., Kern, S. E., Preisinger, A. C., Leppert, M., Nakamura, Y., White, R., Smits, A. M. & Bos, J. L. (1988) Genetic alterations during colorectal-tumor development. *N Engl J Med* 319:525-32.
4. Markowitz, S., Wang, J., Myeroff, L., Parsons, R., Sun, L., Lutterbaugh, J., Fan, R. S., Zborowska, E., Kinzler, K. W., Vogelstein, B., Brattain, M. & Willson, J. K. V. (1995) Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability. *Science* 268:1336-8.
5. Thiagalingam, S. (1996) Evaluation of Chromosome 18q in Colorectal Cancers. *Nature Genetics* 13:343-346.
6. Samuels, Y. & Velculescu, V. E. (2004) Oncogenic mutations of PIK3CA in human cancers. *Cell Cycle* 3:1221-4.
7. Baker, S. J., Fearon, E. R., Nigro, J. M., Hamilton, S. R., Preisinger, A. C., Jessup, J. M., vanTuinen, P., Ledbetter, D. H., Barker, D. F., Nakamura, Y., White, R. & Vogelstein, B. (1989) Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas. *Science* 244:217-21.
8. Joyce, D. L., Wahl, R. L., Patel, P. V., Schulick, R. D., Gearhart, S. L. & Choti, M. A. (2006) Preoperative positron emission tomography to evaluate potentially resectable hepatic colorectal metastases. *Arch Surg* 141:1220-6; discussion 1227.
9. O'Connell, J. B., Maggard, M. A. & Ko, C. Y. (2004) Colon cancer survival rates with the new American Joint Committee on Cancer sixth edition staging. *J Natl Cancer Inst* 96:1420-5.
10. Tanaka, K., Shimada, H., Miura, M., Fujii, Y., Yamaguchi, S., Endo, I., Sekido, H., Togo, S. & Ike, H. (2004) Metastatic tumor doubling time: most important prehepatectomy predictor of survival and nonrecurrence of hepatic colorectal cancer metastasis. *World J Surg* 28:263-70.
11. Muto, T., Bussey, H. J. & Morson, B. C. (1975) The evolution of cancer of the colon and rectum. *Cancer* 36:2251-70.
12. Welin, S., Youker, J. & Spratt, J. S., Jr. (1963) The Rates and Patterns of Growth of 375 Tumors of the Large Intestine and Rectum Observed Serially by Double Contrast Enema Study (Malmoe Technique). *Am J Roentgenol Radium Ther Nucl Med* 90:673-87.
13. Sjoblom, T., Jones, S., Wood, L. D., Parsons, D. W., Lin, J., Barber, T. D., Mandelker, D., Leary, R. J., Ptak, J., Silliman, N., Szabo, S., Buckhaults, P., Farrell, C., Meeh, P., Markowitz, S. D., Willis, J., Dawson, D., Willson, J. K., Gazdar, A. F., Hartigan, J., Wu, L., Liu, C., Parmigiani, G., Park, B. H., Bachman, K. E., Papadopoulos, N., Vogelstein, B., Kinzler, K. W. & Velculescu, V. E. (2006) The consensus coding sequences of human breast and colorectal cancers. *Science* 314:268-74.
14. Wood, L. D., Parsons, D. W., Jones, S., Lin, J., Sjoblom, T., Leary, R. J., Shen, D., Boca, S. M., Barber, T., Ptak, J., Silliman, N., Szabo, S., Dezso, Z., Ustyanksky, V., Nikolskaya, T., Nikolsky, Y., Karchin, R., Wilson, P. A., Kaminker, J. S., Zhang, Z., Croshaw, R., Willis, J., Dawson, D., Shipitsin, M., Willson, J. K., Sukumar, S., Polyak, K., Park, B. H., Pethiyagoda, C. L., Pant, P. V., Ballinger, D. G., Sparks, A. B., Hartigan, J., Smith, D. R., Suh, E., Papadopoulos, N., Buckhaults, P., Markowitz, S.

- D., Parmigiani, G., Kinzler, K. W., Velculescu, V. E. & Vogelstein, B. (2007) The genomic landscapes of human breast and colorectal cancers. *Science* 318:1108-13.
15. Terry, N. H. & White, R. A. (2006) Flow cytometry after bromodeoxyuridine labeling to measure S and G2+M phase durations plus doubling times in vitro and in vivo. *Nat Protoc* 1:859-69.
16. Michel, P., Hemet, J., Paresy, M., Menard, J. F., Laquerriere, A., Scotte, M., Paillot, B., Peillon, C. & Ducrotte, P. (1997) Comparison between endoscopic and surgical sampling for the measurement of potential doubling time in colorectal cancer. *Cytometry* 29:273-8.
17. Schiepers, C., Haustermans, K., Geboes, K., Filez, L., Bormans, G. & Penninckx, F. (1999) The effect of preoperative radiation therapy on glucose utilization and cell kinetics in patients with primary rectal carcinoma. *Cancer* 85:803-11.
18. Terry, N. H., Meistrich, M. L., Rouben, L. D., Lynch, P. M., Dubrow, R. A. & Rich, T. A. (1995) Cellular kinetics in rectal cancer. *Br J Cancer* 72:435-41.
19. Rew, D. A., Wilson, G. D., Taylor, I. & Weaver, P. C. (1991) Proliferation characteristics of human colorectal carcinomas measured in vivo. *Br J Surg* 78:60-6.
20. Fuller, C. E., Davies, R. P., Williams, G. T. & Williams, E. D. (1990) Crypt restricted heterogeneity of goblet cell mucus glycoprotein in histologically normal human colonic mucosa: a potential marker of somatic mutation. *Br J Cancer* 61:382-4.
21. Okayasu, I., Hana, K., Tsuruta, T., Okamura, N., Ogawa, T., Tokuyama, W., Kajita, S., Yoshida, T. & Mikami, T. (2006) Significant increase of colonic mutated crypts correlates with age in sporadic cancer and diverticulosis cases, with higher frequency in the left- than right-side colorectum. *Cancer Sci* 97:362-7.
22. Loeb, L. A., Loeb, K. R. & Anderson, J. P. (2003) Multiple mutations and cancer. *Proc Natl Acad Sci U S A* 100:776-81.
23. Araten, D. J., Golde, D. W., Zhang, R. H., Thaler, H. T., Gargiulo, L., Notaro, R. & Luzzatto, L. (2005) A quantitative measurement of the human somatic mutation rate. *Cancer Res* 65:8111-7.
24. DeMars, R. & Held, K. R. (1972) The spontaneous azaguanine-resistant mutants of diploid human fibroblasts. *Humangenetik* 16:87-110.
25. Elmore, E., Kakunaga, T. & Barrett, J. C. (1983) Comparison of spontaneous mutation rates of normal and chemically transformed human skin fibroblasts. *Cancer Res* 43:1650-5.
26. Baker, S. J., Preisinger, A. C., Jessup, J. M., Paraskeva, C., Markowitz, S., Willson, J. K., Hamilton, S. & Vogelstein, B. (1990) p53 gene mutations occur in combination with 17p allelic deletions as late events in colorectal tumorigenesis. *Cancer Res* 50:7717-22.
27. Dressman, D., Yan, H., Traverso, G., Kinzler, K. W. & Vogelstein, B. (2003) Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proc Natl Acad Sci U S A* 100:8817-22.
28. Diehl, F., Li, M., He, Y., Kinzler, K. W., Vogelstein, B. & Dressman, D. (2006) BEAMing: single-molecule PCR on microparticles in water-in-oil emulsions. *Nat Methods* 3:551-9.
29. Nowak, M. A. (2006) *Evolutionary Dynamics* (Belknap Press).
30. Shibata, D. & Tavaré, S. (2006) Counting divisions in a human somatic cell tree: how, what and why? *Cell Cycle* 5:610-4.
31. Baguley, B. C. & Ferguson, L. R. (1998) Mutagenic properties of topoisomerase-targeted drugs. *Biochim Biophys Acta* 1400:213-22.

32. Koga, H., Moriya, Y., Akasu, T. & Fujita, S. (1999) The relationship between prognosis and CEA-dt after hepatic resection in patients with colorectal carcinomas. *Eur J Surg Oncol* 25:292-6.
33. Staab, H. J., Anderer, F. A., Hornung, A., Stumpf, E. & Fischer, R. (1982) Doubling time of circulating cea and its relation to survival of patients with recurrent colorectal cancer. *Br J Cancer* 46:773-81.
34. Ptashne, M. (2007) On the use of the word 'epigenetic'. *Curr Biol* 17:R233-6.
35. Talmadge, J. E. (2007) Clonal selection of metastasis within the life history of a tumor. *Cancer Res* 67:11471-5.
36. Fidler, I. J. (1989) Origin and biology of cancer metastasis. *Cytometry* 10:673-80.
37. Waghorne, C., Thomas, M., Lagarde, A., Kerbel, R. S. & Breitman, M. L. (1988) Genetic evidence for progressive selection and overgrowth of primary tumors by metastatic cell subpopulations. *Cancer Res* 48:6109-14.
38. Price, J. E., Bell, C. & Frost, P. (1990) The use of a genotypic marker to demonstrate clonal dominance during the growth and metastasis of a human breast carcinoma in nude mice. *Int J Cancer* 45:968-71.
39. Scheel, C., Onder, T., Karnoub, A. & Weinberg, R. A. (2007) Adaptation versus selection: the origins of metastatic behavior. *Cancer Res* 67:11476-9; discussion 11479-80.

Figure Legends

Fig. 1. Major genetic alterations associated with colorectal tumorigenesis. See SI *Text* for further explanation.

Fig. 2. Representative examples of sequencing chromatograms of DNA from a xenograft, from the metastatic lesion from which the xenograft was derived, and from the patient's normal cells. Note that the ratio of the mutant to wild-type allele in the xenograft is higher than that in the metastatic lesion because the latter represented a mixture of neoplastic and non-neoplastic cells (stroma, white blood cells, etc.). The arrow points to the mutated base.

Fig. 3. Histopathology of representative lesions. (A) Primary invasive moderately differentiated adenocarcinoma (black circle) with adjacent non-neoplastic colonic mucosa (red circle) from patient 2. (B) Metastatic adenocarcinoma (black circle) to liver (red circle) derived from primary colon adenocarcinoma of patient 2. (C) Primary invasive moderately differentiated adenocarcinoma (black circle) arising in a tubular adenoma (red circle) from patient 10. All sections were stained with H&E and the tissues within each circle were separately microdissected.

Fig. 4. Representative examples of BEAMing assays from the indicated patients and lesions. In patient 13, the mutation shown represents one which was present in a new metastasis that occurred 29 months following chemotherapy (see text). The red dots correspond to beads attached to mutant DNA fragments (labeled with phycoerythrin [PE]), the blue dots correspond to beads attached to wt DNA fragments (labeled with fluorescein [FITC]) and the black dots correspond to beads attached to both wt and mutant DNA fragments.

Fig. 5. Evolution of a lethal cancer. Each cell-filled cone represents one or more clonal expansions (see SI *Text* for details). The times required for the evolution of the large adenoma founder cell to an advanced carcinoma founder cell ($\Delta_{\text{LAd,ACa}}$), and evolution of the advanced carcinoma founder cell to metastatic founder cell ($\Delta_{\text{ACa,Met}}$) were determined by comparative lesion sequencing. Other intervals, such as the time (T_{exp}) required for the expansion of the metastasis founder cell $\text{FCell}_{\text{Met}}$ to the size detected in our patients, were estimated as described in SI *Text*. The model posits that there are at least two clonal expansions, denoted by question marks, that are not associated with any known genetic alterations.

Table 1. Summary of patient information

Patient #	Wood <i>et al.</i> ID#	Age at Diagnosis	Sex	Location of colorectal tumor	TNM Stage*	Site of index lesion	# mutations in colorectal adenoma / # in carcinoma**	# mutations in colorectal carcinoma / # in index metastasis**	# Other metastases	# mutations in other metastases/ # index metastasis**
1	Mx27	73	F	Ascending	T3N1M0	liver	NA	47/47	3	24/24
2	Mx29	50	M	Descending	T4N1M1	liver	NA	7/7	3	17/17
3	Mx34	83	F	Cecum	T4N2M1	lymph node	17/22	24/25	4	31/31
4	Mx40	75	F	Cecum	T4N1M0	lymph node	NA	5/5	3	9/9
5	Mx43	72	M	Sigmoid	T3N2M1	liver	NA	48/50	5	98/98
6	Co92	47	F	Cecum	T3N2M0	liver	NA	8/8	0	NA
7	Mx32	55	F	Ascending	T3N1M0	liver	NA	28/32	3	39/45
8	Co84	41	M	Cecum	T4N2M1	lymph node	NA	4/4	0	NA
9	Mx38	65	M	Rectum	yT3N1M0	liver	NA	6/6	3	17/17
10	Co82	80	F	Cecum	T3N1M0	colon	6/11	NA	1	5/5
11	Mx26	46	F	Cecum	T2N2M1	liver	NA	NA	1	3/3
12	Co108	76	F	Ascending	T4N0M1	liver	NA	NA	1	6/6
13	Mx41	55	M	Ascending	T3N1M1	liver	NA	49/49***	2	6/6***
Average or Total		63					23/33	226/233	29	255/261

* T2 - carcinoma invaded muscularis propria; T3 - carcinoma invaded through muscularis propria into submucosa; T4 - carcinoma invaded through wall of colon into nearby tissues or organs; N0 - no lymph node involvement; N1 - cancer cells found in 1 to 3 nearby lymph nodes; N2 - cancer found in more than 3 nearby lymph nodes; M0 - No distant metastases identified; M1 - Distant metastasis identified; a "y" prior to the TNM stage means that the patient was treated with chemoradiotherapy prior to surgery to reduce the size of the lesion.

** The numbers refer to the mutations that could be successfully sequenced. Not all mutations in an index metastatic lesion could be sequenced in other lesions of the same patient because of limitations in available material.

*** There were 49 mutations detected in the liver and two lymph node metastases that were removed at the time of surgery. A new metastasis developed 29 months later, following chemotherapy. This late metastasis contained 19 new mutations that were not present in the original metastases or carcinoma and are not included in this Table (see text).

NA, not applicable because indicated comparison could not be performed.

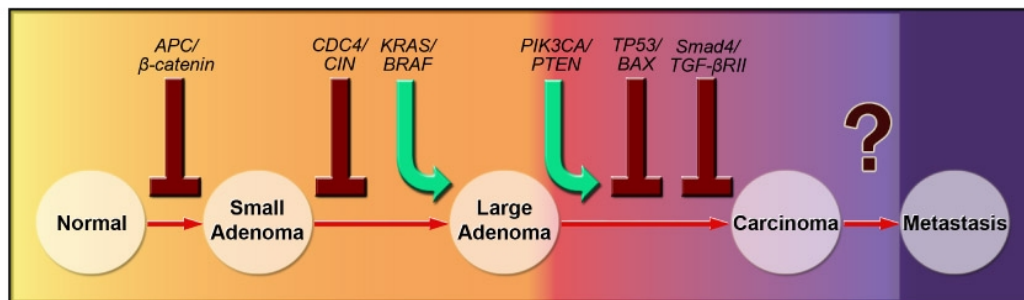
Table 2. Genes specifically mutated in metastases or carcinomas.

A. Mutations identified in metastasis but not precursor colorectal carcinoma

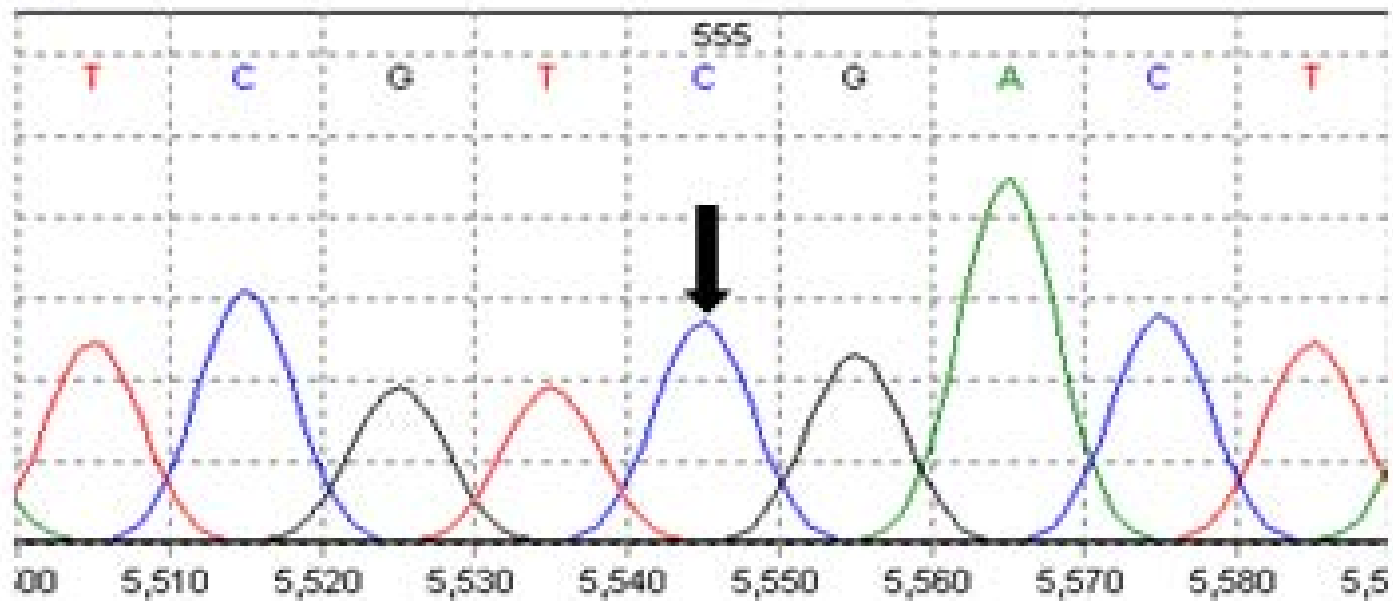
Gene	Patient #
PLCG2	3
CORO1B	5
KCNC4	5
CRB13	7
ENPP2	7
GPR50	7
P2RY14	7

B. Mutations identified in colorectal carcinoma but not precursor adenoma

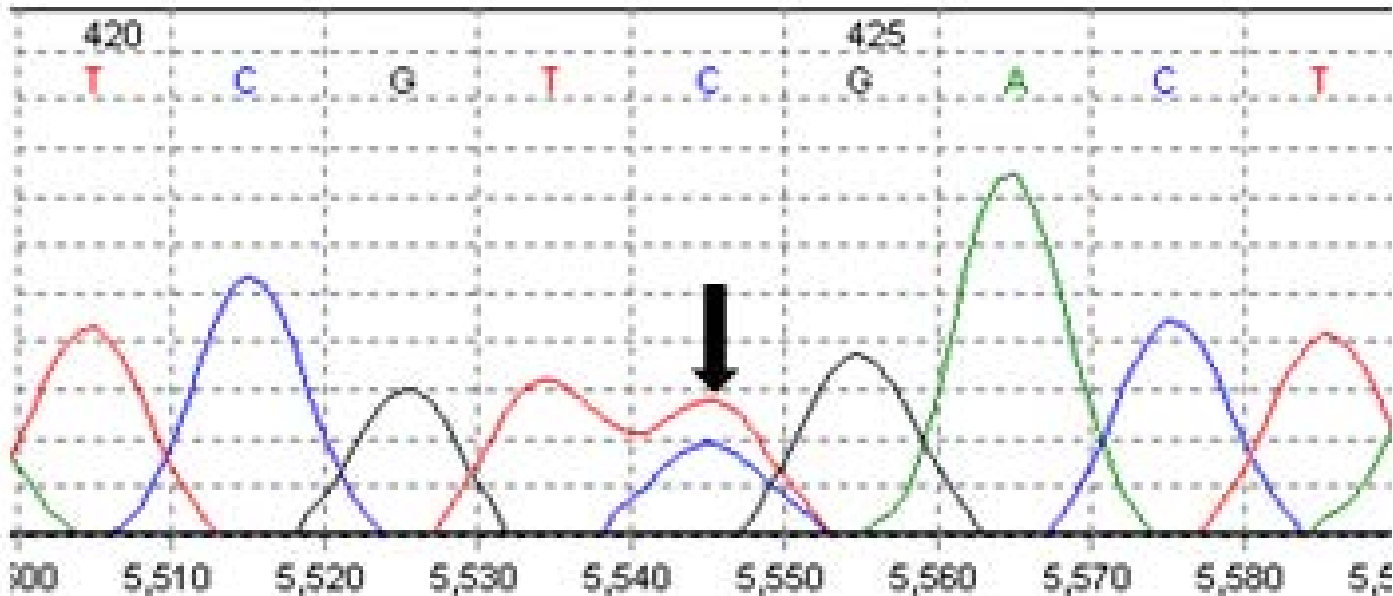
Gene	Patient #
CACNA2D3	3
HUWE1	3
SFRS6	3
TP53	3
RYR2	3
C1QR1	10
AGC1	10
NUP210	10
TP53	10
TTN	10



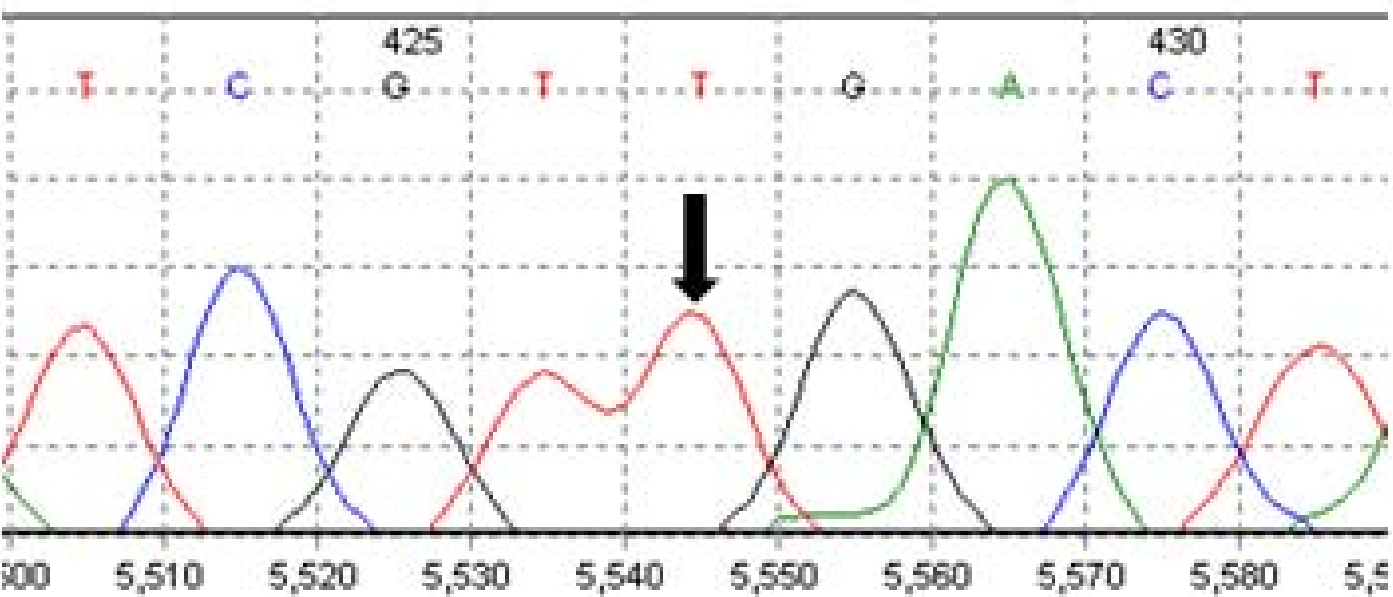
Xenograft

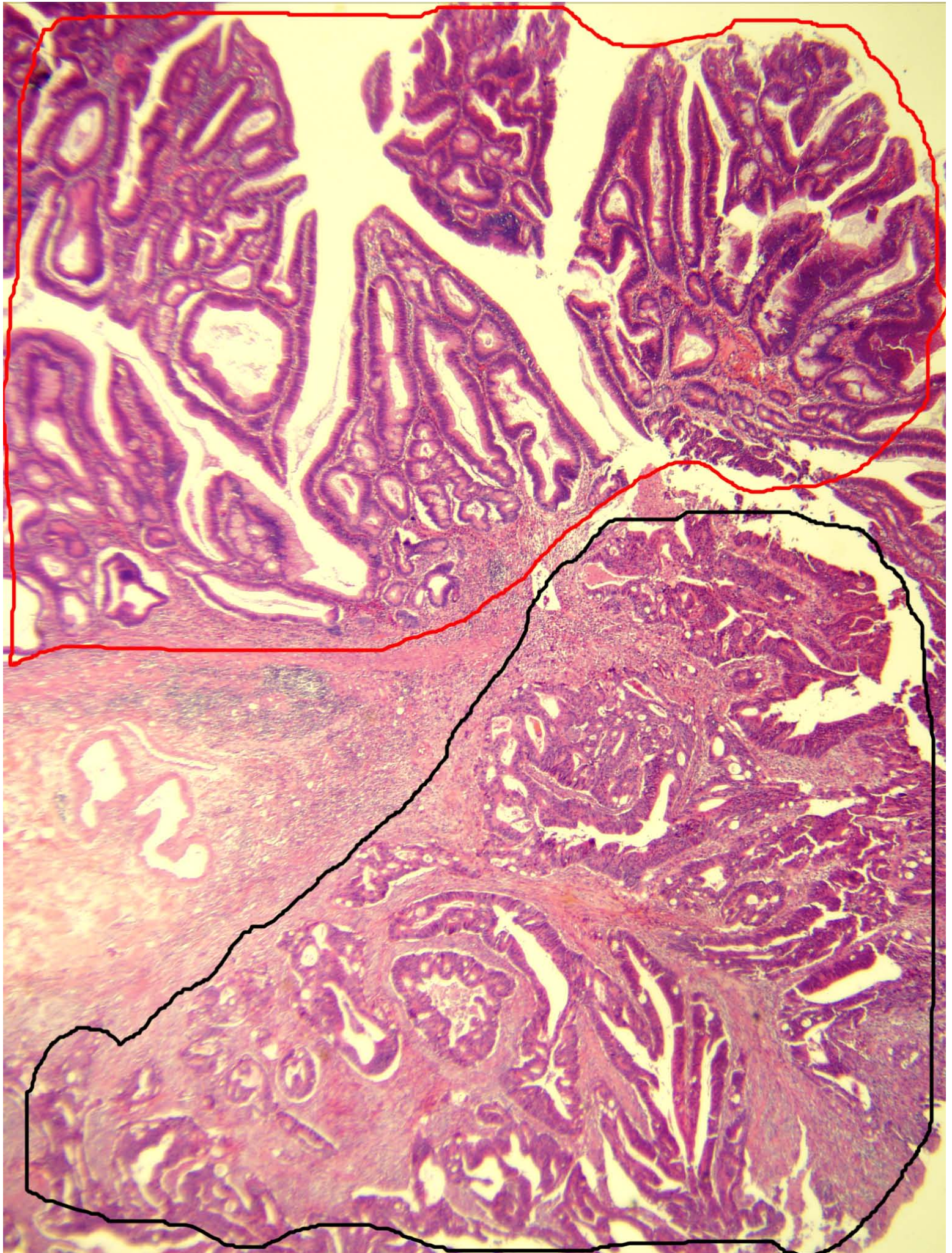


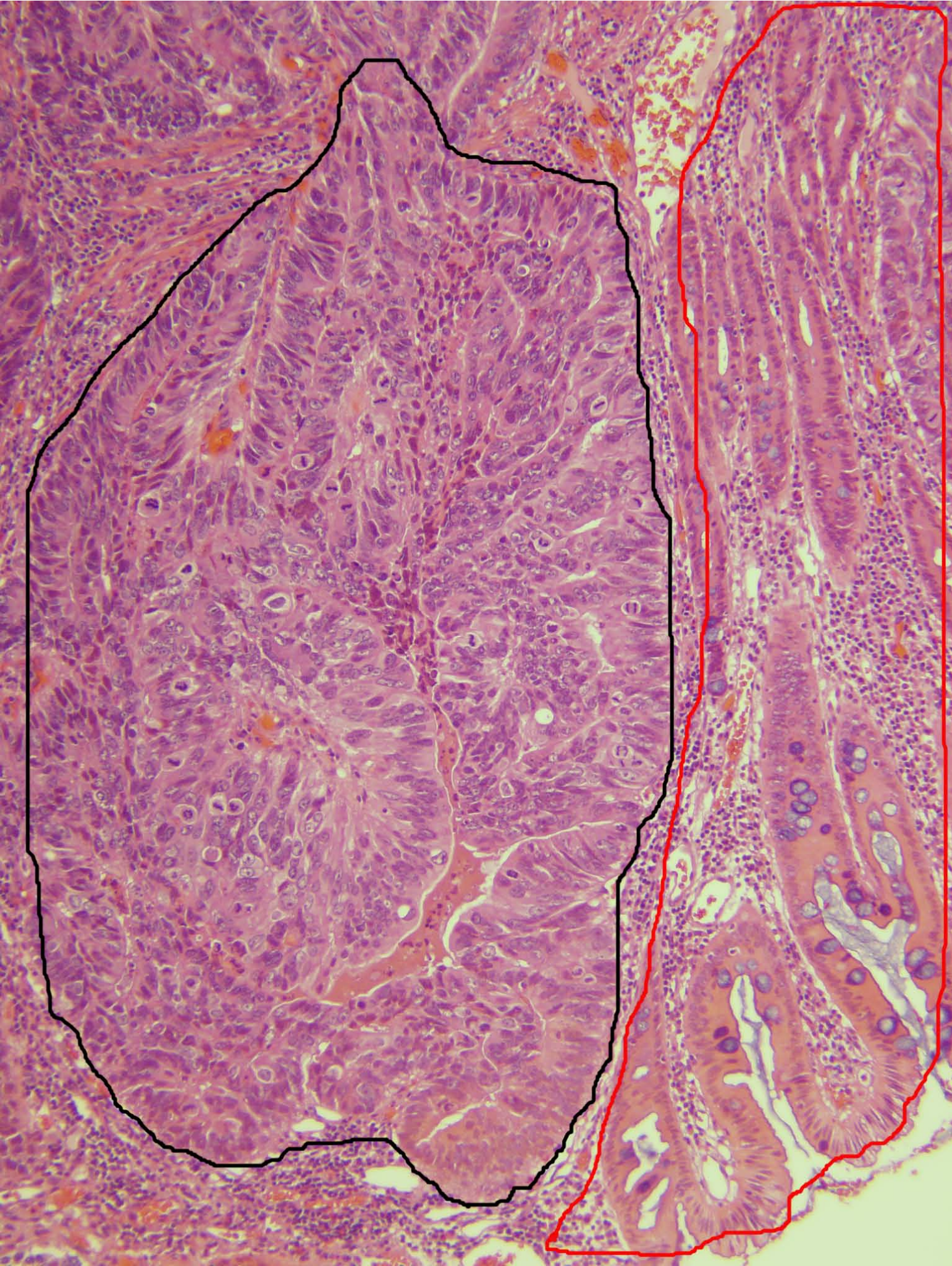
Metastasis

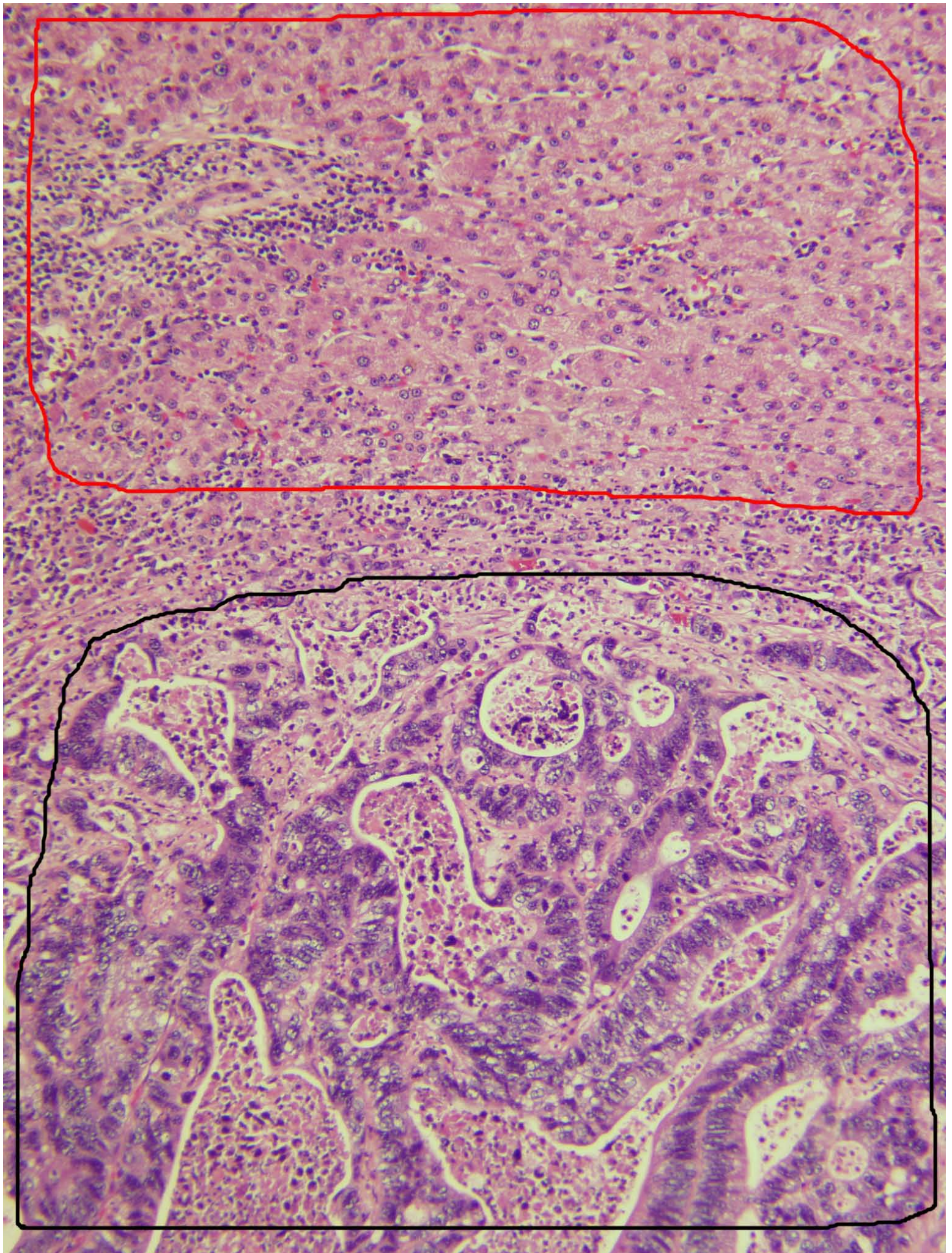


Normal

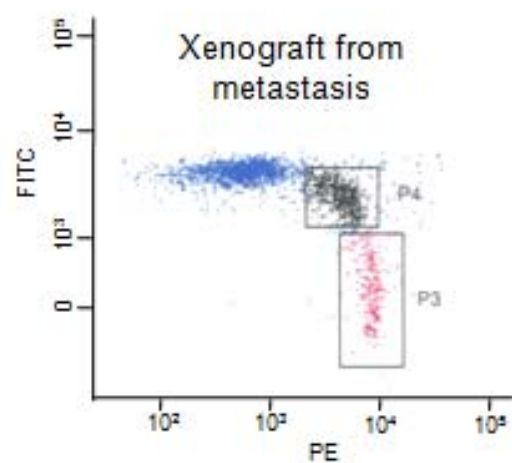




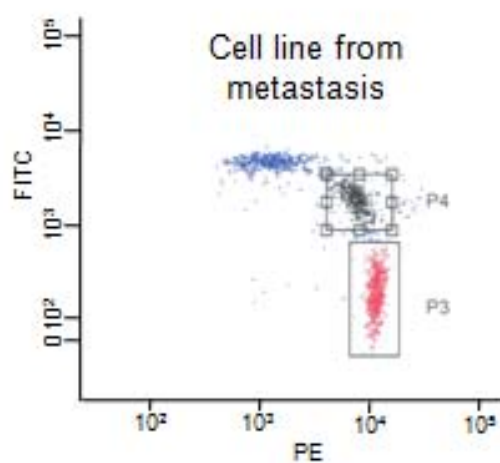




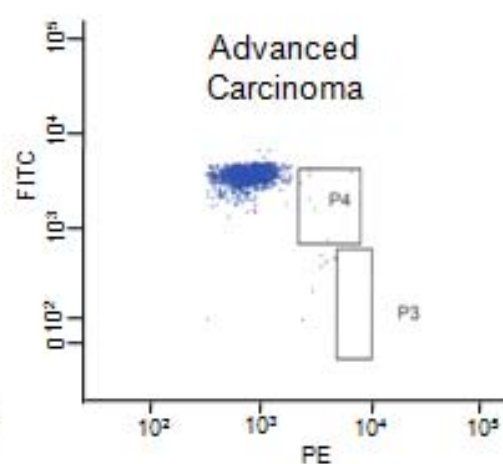
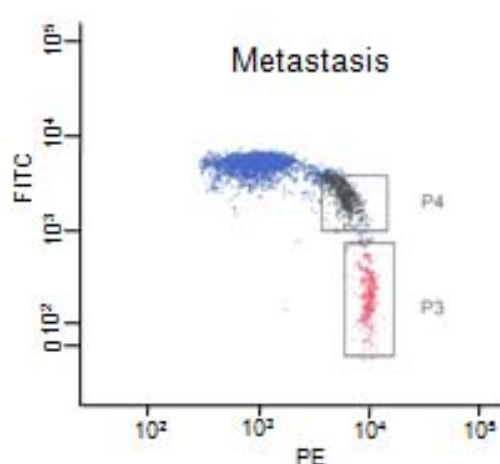
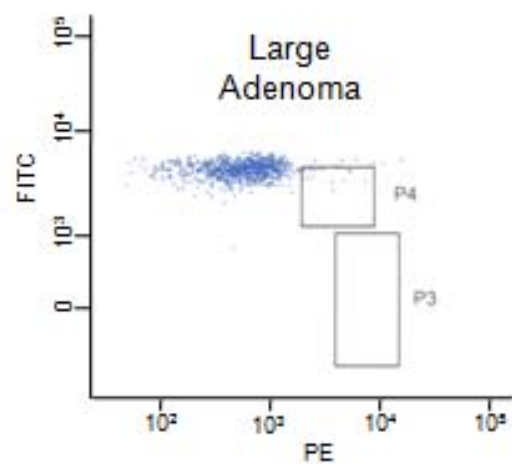
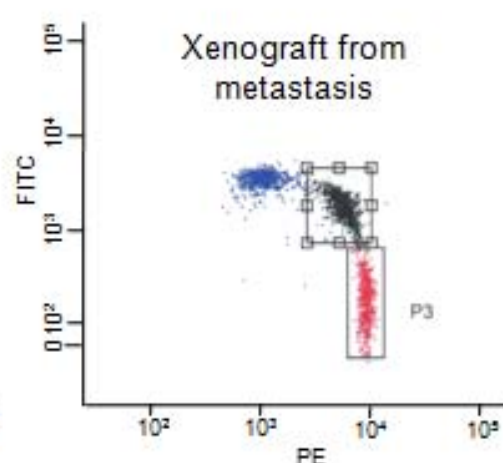
Patient 3

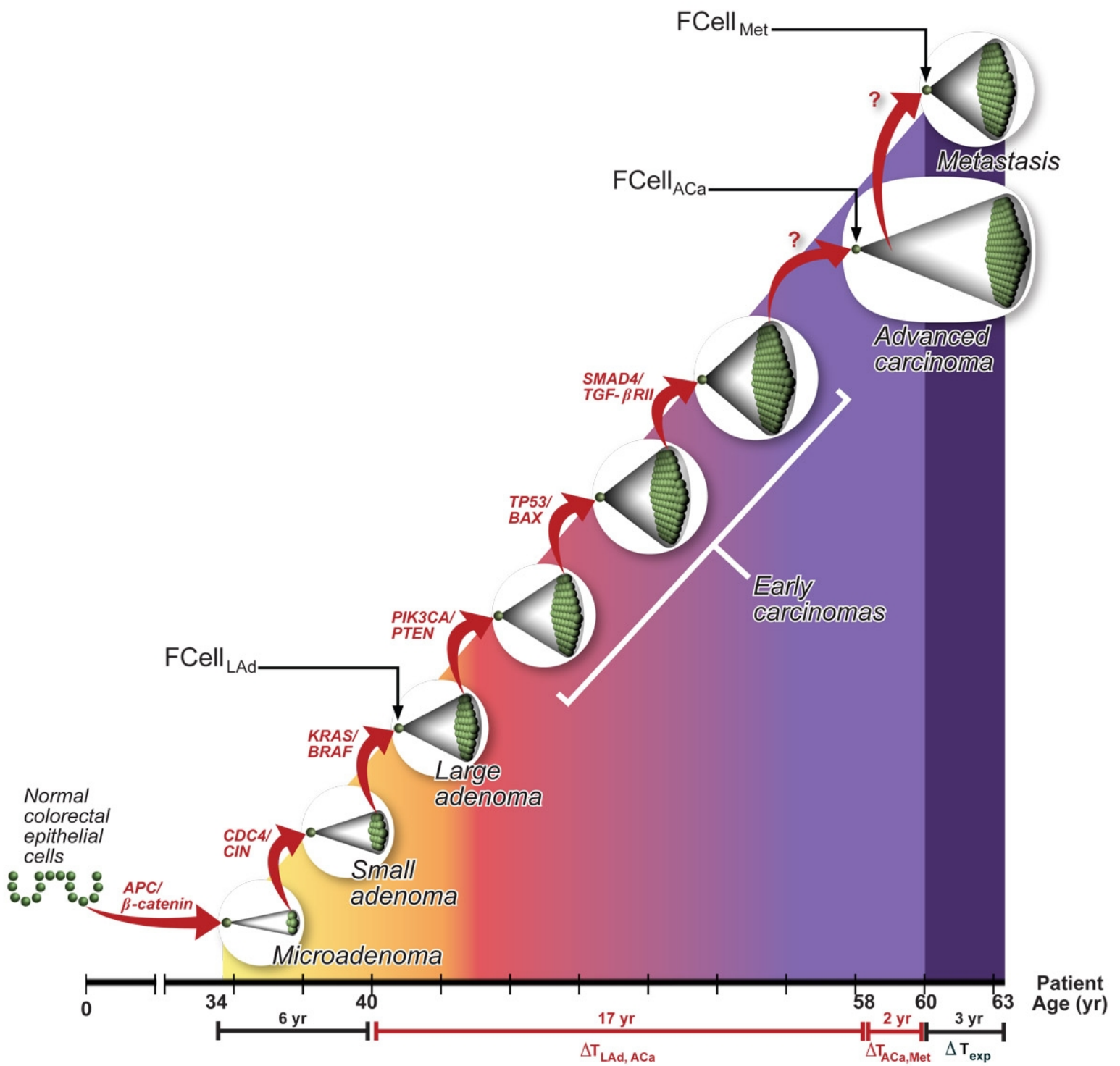


Patient 12



Patient 13





Methods

Samples and DNA purification. DNA samples from tumor samples or their derived xenografts or cell lines were obtained and purified as described previously (1). DNA from frozen tissues was purified by a slight modification of a previously described method (2). In brief, frozen tissue was dissolved in guanidine. After overnight ultracentrifugation through cesium chloride, the RNA was pelleted by centrifugation, the guanidine layer was removed and ~4 ml of the cesium chloride layer containing the genomic DNA was carefully collected. The DNA was precipitated with ethanol, then dissolved in water and treated overnight with proteinase K. The proteinase K-digested DNA was further purified by phenol/chloroform extraction and ethanol precipitation. Enriched populations of neoplastic cells were obtained from paraffin embedded sections by microdissection. DNA was purified from these cells using the QIAamp® DNA Micro Kit from Qiagen (Valencia, CA) as directed by the manufacturer. All samples were obtained in accordance with the Health Insurance Portability and Accountability Act (HIPPA).

DNA Sequencing. PCR amplification and sequencing were carried out as described previously (1). In brief, 289 exons in which a mutation had been identified previously in an index lesion studied in (1) or (3) were PCR-amplified in all other available DNA samples from the patient. DNA samples from xenografts, cell lines and frozen tissues were amplified using the primers described previously (3). New amplicons of smaller size were designed for the DNA purified from paraffin-embedded samples. PCR products >250 bp in length were purified using AMPure (Agencourt Biosciences, Beverly, MA) and sequenced with M13 forward primer (5'-GTAAAACGACGGCCAGT-3') and Big Dye Terminator Kit v.3.1. (Applied Biosystems, Foster City, CA). Smaller amplicons from paraffin-embedded tissues were purified using a DirectPrep 96 Miniprep Kit (Qiagen) and sequenced using Big Dye Terminator kit v.1.1. The CleanSeq kit (Agencourt Biosciences, Beverly, MA) was used to remove excess dye terminators from the reaction before visualization by capillary electrophoresis on an ABI PRISM 3730xl instruments (Applied Biosystems). Sequence traces were aligned to the genomic reference sequence and analyzed using Mutation Surveyor software (SoftGenetics, State College, PA).

Quantification Of Mutation Frequencies In Specific Samples. When sequencing chromatograms were difficult to interpret in the DNA purified from tumor samples, we re-evaluated the mutation in question either by cloning the PCR product and sequencing individual clones or by performing a BEAMing assay. Cloning of PCR products was carried out as previously described (4). For BEAMing, PCR is performed in water-in-oil microemulsions containing beads, specific primers, and Taq polymerase (5). In aqueous compartments containing a template molecule and a bead, the amplification products become tightly bound to the bead. At the end of the process, these beads each contain hundreds of thousands of identical copies of the template molecule. The beads can be analyzed via flow cytometry following hybridization to labeled probes containing mutant or wt sequences. BEAMing assays in the current study were performed essentially as described previously (6). The target region was amplified using gene-specific primers with 5' tags

(F tag: 5'-TCCCGCGAAATTAATACGAC-3'; R tag: 5'-GCTGGAGCTCTGCAGCTA-3') and Phusion high fidelity DNA polymerase (New England Biolabs, Beverly, MA). The resulting products were quantified with the PicoGreen dsDNA assay kit (Invitrogen, Carlsbad, CA). Fifty to 150 pmoles PCR product were used as a templates for emulsion PCR using

F tag-labelled streptavidin-coated beads and F tag and R tag as forward and reverse primers, respectively. The emulsions were broken and beads purified using a magnet. The beads were hybridized to a fluorescein-labeled probe specific for the wt sequence and a biotin-labeled probe specific for the mutant sequence. After incubation with phycoerythrin-labeled streptavidin, mutant and wild type alleles were distinguished by flow cytometry. In addition to using BEAMing to help interpret the result of questionable sequencing chromatograms, we used this technique to determine the maximum number of cells in a tumor cell population that harbored a specific mutation (see main text).

Supplementary Text

Mathematical Analysis of Comparative Lesion Sequencing. The number of somatic mutations (N_i) that have accumulated in cell_{*i*} is linearly related to the mutation rate (μ) by the following equation:

$$N_i = \mu \cdot Q \cdot G_i, \quad [3]$$

where Q is the number of bp in the genome and G_i is the number of generations that the cell has undergone. The mutation rate μ is measured as mutations/bp/generation and is assumed to be constant over time.

If two cells (Cell₁ and Cell₂) have always had the same mutation rate, then the ratio of their accumulated mutations is thus given by

$$N_1/N_2 = G_1/G_2. \quad [4]$$

In the cases considered in this study, founder cell FCell₂ is a progeny of founder cell FCell₁ and both founder cells were derived from a precursor cell that contained no somatic mutations. Hence,

$$N_2 = N_1 + F_{1,2} \cdot N_2, \quad [5]$$

where $F_{1,2}$ is defined as the fraction of the total somatic mutations present in Fcell₂ that are not present in Fcell₁ (i.e., $F_{1,2} = 1 - N_1/N_2$). Note that because Fcell₂ is a progeny of Fcell₁, there can be no mutations in Fcell₁ that are not also in Fcell₂, so $N_2 > N_1$ and $F_{1,2} \geq 0$.

T_i is defined as the time during which N_i somatic mutations have accumulated in cell_{*i*}. T_i is then the product of G_i and the cell cycle time, T_{pot} , i.e.,

$$T_i = G_i \cdot (T_{\text{pot}})_i \quad [6]$$

If the average T_{pot} during mutation accumulation in both cells is identical, then [4] can be reduced to

$$N_1/N_2 = T_1/T_2. \quad [7]$$

Combining [5] and [7] gives

$$\Delta T_{1,2} = T_2 - T_1 = T_2 \cdot F_{1,2} \quad [8]$$

Note that this equation is independent of the actual mutation rate μ or the actual cell cycling time T_{pot} of the tumors that are analyzed. The average ΔT values reported in the text are therefore independent of variations in these two parameters among different

patients or tumors. The accuracy of the estimates of ΔT are, however, dependent on the number of somatic mutations identified, as described in the next section.

Confidence Intervals For Estimates of $\Delta T_{1,2}$. Consider the case when N_1 mutations are observed at T_1 and $N_2 - N_1$ additional mutations are observed at T_2 . Here T_2 is known (say the age at diagnosis) while T_1 is unknown. Mutations are assumed to occur at a constant rate μ and events are assumed to be Poisson distributed. Using prior distributions Beta (a, b) on the unknown T_1 / T_2 , and Gamma ($0, 0$) on μ , we can derive the a posteriori distribution. The Beta and Gamma prior distributions were chosen because of their computational convenience and the interpretability of the input values as events in hypothetical previous experiments (7). Whenever feasible, input values were chosen to represent vague a priori knowledge. The two unknowns turn out to be statistically independent: μ is distributed as Gamma (N_2, T_2) while the ratio T_1 / T_2 is distributed as Beta ($N_1 + a, N_2 - N_1 + b$). Thus $T_2 \cdot (N_2 + a) / (N_1 + a + b)$ is the a posteriori mean of the unknown T_1 given T_2 , and the Beta distribution can be used to construct highest posterior density regions (denoted as 90% CI's in the text). These results apply at the individual patient level. As long as the ratio T_1 / T_2 is constant across patients, the distribution of T_1 / T_2 remains as above, with N_1 and N_2 representing the sum of the mutations found in all patients. It can be proven that this applies also to the case in which each patient's tumor has a different mutation rate. Using this method one can derive the distribution of a birth date T_1 conditional on the next birth date T_2 . We applied this recursively to determine earlier birthdates such as T_{LAd} , in which case uncertainty about T_{ACa} is also taken into account.

For pooled analyses and for the transition from T_{LAd} to T_{ACa} we used "noninformative" choices $a=b=0$. For patient-specific analysis of transitions from T_{ACa} to T_{Met} this was not possible because in some cases there were no mutations that were present in the metastasis but absent in the primary colorectal carcinoma. We therefore used an empirical Bayes approach for T_{ACa} to T_{Met} and estimated $(a+b)$ as $(1/\phi) - 1$ where ϕ is the "overdispersion parameter" (8) for a beta-binomial model. This approximation allows for some variation in T_1 / T_2 across patients. Also, we took $a/(a+b)$ to be the relevant overall proportion across patients.

All statistical analyses were performed in the statistical package R. The 'aod' library was used to estimate the beta-binomial model.

Assumptions and Other Areas of Uncertainty.

(i) *Mutation reversibility.* One of the assumptions made in deriving Equation 1 is that mutations are irreversible and that all mutations observed in F_{cell1} are present in F_{cell2} . We know of no evidence inconsistent with this assumption. Moreover, it is supported by the fact that nearly every mutation initially discovered in a late lesion (e.g., metastasis), then found in the carcinoma of the same patient, was also found in all other metastatic lesions from that patient (247 of 248 instances examined).

(ii) *Mutation constancy.* We assume that the mutation rate during normal epithelial stem cell growth and tumorigenesis is constant. This is not true for tumors with mutator phenotypes, such as those with mismatch repair deficiency, as the rate of mutations in these tumors increases by 100-fold or more in the tumors once both alleles

of the MMR gene are inactivated (9, 10). However, none of the tumors evaluated in the current study, or in that of (3) were MMR-deficient. Moreover, as described in the main text, the mutation rate measured in the analyzed tumors was similar to, but slightly less, than those measured in normal cells. Somatic mutation rate measurements in normal cells such as fibroblasts or lymphoblasts have yielded highly consistent results. For example, mutations in HGPRT or glycophorin A result in dominant phenotypes that have been measured in several studies as $\sim 1 \times 10^{-6}$ mutations/gene/generation (11-13). Given the conventional rule of thumb that there are ~ 1000 nt in or around these genes that can result in the mutant phenotype (14), this is equivalent to $\sim 10 \times 10^{-10}$ mutations/bp/generation. Estimates of the *in vivo* mutation rate of human colorectal epithelial stem cells can be made from the study of O-acetylated sialoglycoproteins (15, 16). Such studies have shown a gradual accumulation of mutations with age. Based on the published correlation of age vs. mutation frequency, we calculate a rate of $\sim 3.2 \times 10^{-10}$ mutations/bp/generation after applying the 1000 nt convention and the T_{pot} described in section (iv) below. This value is similar to the one we calculated for cancer cells ($\sim 4.6 \times 10^{-10}$ mutations/bp/generation). These results, in aggregate, suggest that the rates of mutation of colorectal epithelial cells, whether normal or neoplastic, are very similar. Though this rate may differ between patients, such inter-patient differences are immaterial to our analysis: the assumption made in the current study is that the mutation rate does not change over time within any given individual. The average somatic mutation rate in tumors in the main text was calculated according to [3], using N of 847 mutations, Q of 340 Mb, and G of 5384 generations ($G = \text{average } T_{Met} \text{ divided by } T_{pot}$).

(iii) *The nature of clonal mutations.* "Clonal mutations" are defined as those which are present in every cell of an analyzed population. By definition, each of these clonal mutations are present in the founder cell of the population. As the progeny of this founder cell increase in number, additional mutations (sub-clonal) accumulate. But such sub-clonal mutations are not relevant to our model.

The notion of clonality and its relationship to founder cells can be confusing but is essential for understanding the derivations of equations [1] to [8]. For example, the adenoma cells we purified could have had clonal mutations that were not present in the analyzed carcinoma. This could occur if founder cell of the carcinoma branched off early during the evolution of the adenoma, but then the adenoma evolved further, accumulating additional mutations that allowed it to grow larger but not to grow invasively (i.e., not allowing it to become a carcinoma). The initial adenoma cells that developed from FCell_{Lad} and gave rise to the carcinoma could have been destroyed and replaced by subsequent round(s) of clonal expansion that resulted in the adenoma cells we analyzed. Because this progressed adenomatous lesion will have all of the mutations present in Fcell_{Lad}, however, the analysis and calculated $\Delta T_{LAd,ACa}$ are unaltered.

Though not performed here, one could use the same heuristic approach to determine the evolutionary time separating the founder cells of two lesions when neither founder cell is a direct descendent of the other, i.e., cell populations on different branches of a tumor evolutionary tree. For example, one could analyze a lymph node metastasis and a liver metastasis even when the latter did not develop directly from the

former. To evaluate the time separating the founder cell birth dates of these lesions, one would have to perform an unbiased mutational analysis of each lesion in an independent manner and could not rely on an evaluation of only those mutations that were present in the most advanced lesion (as was done in the current study).

(iv) Estimates of T_{pot} . T_{pot} is defined as the cell division time that would occur in the absence of any cell death. This parameter was used in two instances in our analysis. First, we used a value of T_{pot} of 4 days in the estimation of somatic mutation rates in cancers as described in (ii) above. Studies in hundreds of patients have shown that the value of T_{pot} generally ranges from 3 to 5 days (17, 18) (19-21). Substitution of 3 or 5 days for 4 days as the T_{pot} would not substantially alter the conclusion that the somatic point mutation rate in cancers is very similar to that observed in normal cells. Second, it was assumed that the T_{pot} of normal and neoplastic colorectal epithelial cells is identical throughout life in deriving [7]. This equation is not dependent on the actual value of T_{pot} , but as with the mutation rate, requires that it be constant throughout the lifetime of an individual patient. Justification for this assumption in neoplastic cells is provided by the fact that the measured values of T_{pot} are similar (3 to 5 days) in patients with different stage lesions (17-21). The T_{pot} of normal human colorectal epithelial stem cells has not been measured. However, a value of 4 days for the T_{pot} of such normal cells seems reasonable given the studies of O-acetylated sialoglycoproteins noted above plus the observation that stem cells in the mouse intestine cycle very frequently (periods of one day in small intestine, somewhat longer in large intestine) (22) .

(v) *Estimates of T_2* . In Equation [8], ΔT is defined as the product of T_2 and $F_{1,2}$. $F_{1,2}$ is determined experimentally but T_2 must be estimated. T_2 is defined as the age of the patient when F_{cell_2} was born. When F_{cell_2} is the founder cell of a metastasis,

$$T_2 = T_{\text{Met}} = T_{\text{dx}} - T_{\text{exp}}, \quad [9]$$

where T_{dx} is the age of the patient when the lesion was detected and T_{exp} is the time during which $F_{\text{cell}_{\text{Met}}}$ underwent the clonal expansion that resulted in the metastasis.

Minimum and maximum estimates of T_{exp} can be obtained from previous studies. For example, a metastatic lesion of 3.5 cm contains 2.24×10^{10} cells, assuming 10^9 cells/cm³. Starting from one cell, this expansion represents 34.3 doublings. If there is absolutely no death of cells, so that tumor size is only limited by the cell cycle time T_{pot} (17-20); Rew, 1991 #7303}, then $T_{\text{exp}} = 34.3 \text{ doublings} \times 4 \text{ days/doubling} = 0.38 \text{ years}$. If on the other hand, the measured tumor doubling time (23-25) is used, then $T_{\text{exp}} = 34.3 \text{ generations} \times 60 \text{ days/generation} = 5.6 \text{ years}$. In the current study, we estimated T_{exp} for metastases as 2 years based on the expectation that the cells would multiply rapidly at the outset of the expansion, when nutrients and angiogenesis are not limiting, but would multiply at the rate predicted by the tumor doubling time once they became radiologically visible (26) . Because these doubling times differ so much among patients, we made no attempt to estimate their values in individual patients and instead used 3 years for all patients. Note that whichever of these estimates is used, T_{exp} is less than 10% of T_{dx} , as the average age of our patients was 63 years. From [9], the estimate of T_{exp} will not substantially affect T_{Met} and from [8], the estimate of T_{Met} will not have a large influence on ΔT .

(vi) *Timelines in Fig. 5.* The time required for the clonal expansion giving rise to the metastasis is T_{exp} , as described in section (v) above. The average T_{Met} was calculated using [9]. The average T_{ACa} was calculated as $T_{\text{dx}} - T_{\text{exp}} - \Delta T_{\text{ACa, Met}}$. The average T_{LAd} was calculated using the data in Table 1 and Equation [2]. The time required for growth of an initiated cell to a large adenoma was estimated as $Y - \Delta T_{\text{ACa, Met}} - \Delta T_{\text{IAC, ACa}}$ where Y is the average time thought to be required for development of an advanced carcinoma from an adenoma. We estimated Y as 25 years on the basis of published studies that evaluated patients with sporadic or familial colorectal neoplasms over time (26, 27).

(vii) *Other features of Figures 1 and 5.* It is assumed that the founder cell of the microadenoma was a normal colorectal epithelial stem cell such as that described in Barker *et al.* (22). Whether this stem cell is located at the bottom (as shown) or tops of the crypts is debatable (28). Modeling studies suggest that chromosomal instability (CIN) occurs relatively early in colorectal tumorigenesis ((29)). Recent data shows that *CDC4* and several other putative *CIN* genes affect chromatid cohesion (30). Because *CDC4* mutations are known to occur in small adenomas (31), alterations of the *CDC4/CIN* pathway are placed early in the model. It is known that *KRAS/BRAF* pathway gene mutations occur in large adenomas but rarely in small adenomas. Mutations in *PIK3CA/PTEN*, *TP53/BAX*, or *SMAD4/TGF β RII* pathway genes are rarely observed in large adenomas but are often observed in carcinomas. These three pathways are therefore likely to be involved in the transition from benign lesions (large adenomas) to progressive malignancies (early carcinomas). Though represented as discrete steps in the model, these stages represent a continuum. The relative order of *PIK3CA/PTEN*, *TP53/BAX*, or *SMAD4/TGF β RII* mutations is conjectural, though it is broadly consistent with previous studies (e.g., (32-38)). Though each of the indicated pathways are likely altered in the majority of colorectal cancers, every cancer does not have an alteration in every pathway; variations among tumors contribute to their biologic heterogeneity (1). The genes listed adjacent to the arrows represent the most frequently mutated genes in the pathways; other mutant genes can affect the same pathways (39, 40). Each cone represents one or more clonal expansions, as explained in section (iii) above. It has been estimated that 10 to 20 driver mutations are accumulated during the tumorigenic process (41); some of these are likely to be responsible for additional clonal expansions within the cones. We speculate that at least one additional mutation is required to evolve an advanced carcinoma (stage T3 or T4, defined as having invaded through the muscularis propria; see legend to Table 1) from an early carcinoma. The pathway(s) responsible for the transition from early to advanced carcinoma has not yet been identified, though some evidence suggests that the *SMAD4/TGF β RII* pathway may play a role (34-38). As noted in the main text, it is not clear whether the process of metastasis requires any additional mutations other than those observed in the advanced carcinoma.

SI References

1. Sjoblom, T., Jones, S., Wood, L. D., Parsons, D. W., Lin, J., Barber, T. D., Mandelker, D., Leary, R. J., Ptak, J., Silliman, N., Szabo, S., Buckhaults, P., Farrell, C., Meeh, P., Markowitz, S. D., Willis, J., Dawson, D., Willson, J. K., Gazdar, A. F., Hartigan, J., Wu, L., Liu, C., Parmigiani, G., Park, B. H., Bachman, K. E., Papadopoulos, N., Vogelstein, B., Kinzler, K. W. & Velculescu, V. E. (2006) The consensus coding sequences of human breast and colorectal cancers. *Science* 314:268-74.
2. Davis, L., Kuehl, M. & Battey, J. (1995) *Basic Methods in Molecular biology* (A[;etpm * :amge.
3. Wood, L. D., Parsons, D. W., Jones, S., Lin, J., Sjoblom, T., Leary, R. J., Shen, D., Boca, S. M., Barber, T., Ptak, J., Silliman, N., Szabo, S., Dezso, Z., Ustyanksky, V., Nikolskaya, T., Nikolsky, Y., Karchin, R., Wilson, P. A., Kaminker, J. S., Zhang, Z., Croshaw, R., Willis, J., Dawson, D., Shipitsin, M., Willson, J. K., Sukumar, S., Polyak, K., Park, B. H., Pethiyagoda, C. L., Pant, P. V., Ballinger, D. G., Sparks, A. B., Hartigan, J., Smith, D. R., Suh, E., Papadopoulos, N., Buckhaults, P., Markowitz, S. D., Parmigiani, G., Kinzler, K. W., Velculescu, V. E. & Vogelstein, B. (2007) The genomic landscapes of human breast and colorectal cancers. *Science* 318:1108-13.
4. Chen, W. D., Han, Z. J., Skoletsky, J., Olson, J., Sah, J., Myeroff, L., Platzer, P., Lu, S., Dawson, D., Willis, J., Pretlow, T. P., Lutterbaugh, J., Kasturi, L., Willson, J. K., Rao, J. S., Shuber, A. & Markowitz, S. D. (2005) Detection in fecal DNA of colon cancer-specific methylation of the nonexpressed vimentin gene. *J Natl Cancer Inst* 97:1124-32.
5. Dressman, D., Yan, H., Traverso, G., Kinzler, K. W. & Vogelstein, B. (2003) Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proc Natl Acad Sci U S A* 100:8817-22.
6. Diehl, F., Li, M., He, Y., Kinzler, K. W., Vogelstein, B. & Dressman, D. (2006) BEAMing: single-molecule PCR on microparticles in water-in-oil emulsions. *Nat Methods* 3:551-9.
7. Gelman, A., Carlin, J. B., Stem, H. S. & Rubin, D. B. (2004) *Bayesian data analysis* (Chapman and Hall/CRC, Boca Raton, Florida).
8. Griffiths, D. A. (1973) Maximum likelihood estimation for the beta-binomial distribution and an application to the household distribution of the total number of cases of disease. *Biometrics* 29:637-648.
9. Shibata, D. & Aaltonen, L. A. (2001) Genetic predisposition and somatic diversification in tumor development and progression. *Adv Cancer Res* 80:83-114.
10. Yamamoto, H., Imai, K. & Perucho, M. (2002) Gastrointestinal cancer of the microsatellite mutator phenotype pathway. *J Gastroenterol* 37:153-63.
11. Araten, D. J., Golde, D. W., Zhang, R. H., Thaler, H. T., Gargiulo, L., Notaro, R. & Luzzatto, L. (2005) A quantitative measurement of the human somatic mutation rate. *Cancer Res* 65:8111-7.

12. DeMars, R. & Held, K. R. (1972) The spontaneous azaguanine-resistant mutants of diploid human fibroblasts. *Humangenetik* 16:87-110.
13. Elmore, E., Kakunaga, T. & Barrett, J. C. (1983) Comparison of spontaneous mutation rates of normal and chemically transformed human skin fibroblasts. *Cancer Res* 43:1650-5.
14. Drake, J. W., Charlesworth, B., Charlesworth, D. & Crow, J. F. (1998) Rates of spontaneous mutation. *Genetics* 148:1667-86.
15. Fuller, C. E., Davies, R. P., Williams, G. T. & Williams, E. D. (1990) Crypt restricted heterogeneity of goblet cell mucus glycoprotein in histologically normal human colonic mucosa: a potential marker of somatic mutation. *Br J Cancer* 61:382-4.
16. Okayasu, I., Hana, K., Tsuruta, T., Okamura, N., Ogawa, T., Tokuyama, W., Kajita, S., Yoshida, T. & Mikami, T. (2006) Significant increase of colonic mutated crypts correlates with age in sporadic cancer and diverticulosis cases, with higher frequency in the left- than right-side colorectum. *Cancer Sci* 97:362-7.
17. Michel, P., Hemet, J., Paresy, M., Menard, J. F., Laquerriere, A., Scotte, M., Paillot, B., Peillon, C. & Ducrotte, P. (1997) Comparison between endoscopic and surgical sampling for the measurement of potential doubling time in colorectal cancer. *Cytometry* 29:273-8.
18. Michel, P., Paresy, M., Lepessot, F., Mauillon, J., Seng, S., Scotte, M., Peillon, C., Ducrotte, P., Hemet, J. & Paillot, B. (1999) Effect of radiation therapy on the potential doubling time of tumours in colorectal cancers. *Eur J Gastroenterol Hepatol* 11:309-14.
19. Schiepers, C., Haustermans, K., Geboes, K., Filez, L., Bormans, G. & Penninckx, F. (1999) The effect of preoperative radiation therapy on glucose utilization and cell kinetics in patients with primary rectal carcinoma. *Cancer* 85:803-11.
20. Terry, N. H., Meistrich, M. L., Roubein, L. D., Lynch, P. M., Dubrow, R. A. & Rich, T. A. (1995) Cellular kinetics in rectal cancer. *Br J Cancer* 72:435-41.
21. Rew, D. A., Wilson, G. D., Taylor, I. & Weaver, P. C. (1991) Proliferation characteristics of human colorectal carcinomas measured in vivo. *Br J Surg* 78:60-6.
22. Barker, N., van Es, J. H., Kuipers, J., Kujala, P., van den Born, M., Cozijnsen, M., Haegebarth, A., Korving, J., Begthel, H., Peters, P. J. & Clevers, H. (2007) Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature* 449:1003-7.
23. Koga, H., Moriya, Y., Akasu, T. & Fujita, S. (1999) The relationship between prognosis and CEA-dt after hepatic resection in patients with colorectal carcinomas. *Eur J Surg Oncol* 25:292-6.
24. Staab, H. J., Anderer, F. A., Hornung, A., Stumpf, E. & Fischer, R. (1982) Doubling time of circulating cea and its relation to survival of patients with recurrent colorectal cancer. *Br J Cancer* 46:773-81.
25. Tanaka, K., Shimada, H., Miura, M., Fujii, Y., Yamaguchi, S., Endo, I., Sekido, H., Togo, S. & Ike, H. (2004) Metastatic tumor doubling time: most important prehepatectomy predictor of survival and nonrecurrence of hepatic colorectal cancer metastasis. *World J Surg* 28:263-70.

26. Welin, S., Youker, J. & Spratt, J. S., Jr. (1963) The Rates and Patterns of Growth of 375 Tumors of the Large Intestine and Rectum Observed Serially by Double Contrast Enema Study (Malmoe Technique). *Am J Roentgenol Radium Ther Nucl Med* 90:673-87.
27. Muto, T., Bussey, H. J. & Morson, B. C. (1975) The evolution of cancer of the colon and rectum. *Cancer* 36:2251-70.
28. Shih, I. M., Wang, T. L., Traverso, G., Romans, K., Hamilton, S. R., Ben-Sasson, S., Kinzler, K. W. & Vogelstein, B. (2001) Top-down morphogenesis of colorectal tumors. *Proc Natl Acad Sci U S A* 98:2640-5.
29. Nowak, M. A., Komarova, N. L., Sengupta, A., Jallepalli, P. V., Shih, I. M., Vogelstein, B. & Lengauer, C. (2002) The role of chromosomal instability in tumor initiation. *Proc Natl Acad Sci U S A* 99:16226-31.
30. Barber, T. D., McManus, K. J., Yuen, K. W. Y., Reis, M., Parmigiani, G., Barrett, I., Nouhi, Y., Spencer, F., Markowitz, S., Velculescu, V. E., Kinzler, K. W., Vogelstein, B., Lengauer, C. & Hieter, P. (2008) Chromatid Cohesion Defects May Underlie Chromosome Instability in Human Colorectal Cancers *Submitted*.
31. Rajagopalan, H., Jallepalli, P. V., Rago, C., Velculescu, V. E., Kinzler, K. W., Vogelstein, B. & Lengauer, C. (2004) Inactivation of hCDC4 can cause chromosomal instability. *Nature* 428:77-81.
32. Samuels, Y. & Velculescu, V. E. (2004) Oncogenic mutations of PIK3CA in human cancers. *Cell Cycle* 3:1221-4.
33. Baker, S. J., Preisinger, A. C., Jessup, J. M., Paraskeva, C., Markowitz, S., Willson, J. K., Hamilton, S. & Vogelstein, B. (1990) p53 gene mutations occur in combination with 17p allelic deletions as late events in colorectal tumorigenesis. *Cancer Res* 50:7717-22.
34. Losi, L., Bouzourene, H. & Benhattar, J. (2007) Loss of Smad4 expression predicts liver metastasis in human colorectal cancer. *Oncol Rep* 17:1095-9.
35. Maitra, A., Molberg, K., Albores-Saavedra, J. & Lindberg, G. (2000) Loss of Dpc4 expression in colonic adenocarcinomas correlates with the presence of metastatic disease. *Am J Pathol* 157:1105-11.
36. Miyaki, M., Iijima, T., Konishi, M., Sakai, K., Ishii, A., Yasuno, M., Hishima, T., Koike, M., Shitara, N., Iwama, T., Utsunomiya, J., Kuroki, T. & Mori, T. (1999) Higher frequency of Smad4 gene mutation in human colorectal cancer with distant metastasis. *Oncogene* 18:3098-103.
37. Ohtaki, N., Yamaguchi, A., Goi, T., Fukaya, T., Takeuchi, K., Katayama, K., Hirose, K. & Urano, T. (2001) Somatic alterations of the DPC4 and Madr2 genes in colorectal cancers and relationship to metastasis. *Int J Oncol* 18:265-70.
38. Tarafa, G., Villanueva, A., Farre, L., Rodriguez, J., Musulen, E., Reyes, G., Seminago, R., Olmedo, E., Paules, A. B., Peinado, M. A., Bachs, O. & Capella, G. (2000) DCC and SMAD4 alterations in human colorectal and pancreatic tumor dissemination. *Oncogene* 19:546-55.
39. Vogelstein, B. & Kinzler, K. W. (2004) Cancer genes and the pathways they control. *Nat Med* 10:789-99.
40. Parsons, D. W., Wang, T. L., Samuels, Y., Bardelli, A., Cummins, J. M., DeLong, L., Silliman, N., Ptak, J., Szabo, S., Willson, J. K., Markowitz, S., Kinzler, K. W.,

- Vogelstein, B., Lengauer, C. & Velculescu, V. E. (2005) Colorectal cancer: mutations in a signalling pathway. *Nature* 436:792.
41. Beerenwinkel, N., Antal, T., Dingli, D., Traulsen, A., Kinzler, K. W., Velculescu, V. E., Vogelstein, B. & Nowark, M. A. (2007) Genetic progression and the waiting time to cancer. *PLoS Computational Biology*:in press.